# THE PURIFICATION OF G4 RESOLVASE1:

# A PROTEIN ENTANGLED IN MANY CELLULAR FUNCTIONS

**AN HONORS THESIS (HONORS 499)**

**BY**

**PETER EDWARD BEERBOWER**

**THESIS ADVISOR**

**DR. PHILIP J. SMALDINO**

**BALL STATE UNIVERSITY**

**MUNCIE, INDIANA**

**MAY 2018**

**EXPECTED DATE OF GRADUATION**

**MAY 2018**

## ABSTRACT

G-quadruplexes are supersecondary structures that form in guanine-rich regions of DNA and RNA. These structures are extremely thermally stable, likely to form in ~750,000 locations in the human genome and are enriched in the promoter regions of proto-oncogenes and developmental genes, non-coding regions, and within telomeres. Due to the prevalence of these structures in key genomic locations, these structures are cable of regulating a large fraction of cellular processes. Within humans, the enzyme G4 Resolvase 1 (G4R1) is responsible for the majority of G-quadruplex resolving activity. This places G4R1 at the forefront of regulating all functions involved with G-quadruplexes. Despite regulating G-quadruplexes, a nearly ubiquitous structure throughout the human genome, G4R1 is relatively understudied. Within this thesis, I present a method for producing highly pure and selectively catalytically active samples of rG4R1 and discuss the potential involvement of G4R1 with the pathology of Amyotrophic Lateral Sclerosis.

## ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Dr. Philip Smaldino, my mentor and supervisor, who has helped guide my development as a professional researcher. I have greatly benefitted from his focus on developing a set of scientific knowledge and skills in tandem with a mindset to appropriately implement what I have learned. He is a keen researcher, an inspired scientist, and a devoted teacher.

I would like to thank Dr. Ann Blakey, my academic advisor, for her continued support and guidance.

## TABLE OF CONTENTS

## TABLE OF FIGURES & TABLES

**FIGURES** **Page**



**TABLES** **Page**

## PROCESS ANALYSIS STATEMENT

Beyond just experiments, scientific research is based on the careful reading and analysis of current literature, the usage of current literature and research interests to apply for grants to fund research, and the dissemination of experimental results through poster presentations and paper publications. While working on research for this thesis, I have been able to gain experience in all of these crucial areas. This experience has helped prepare me for my future as a research scientist.

While writing this thesis, I have greatly improved my ability to analyze current scientific literature and search available biological databases for relevant information. These skills have proven invaluable for proposing future research in grants, papers, and in discussions with other scientists.

# CHAPTER ONE

## Purification of Catalytically Active G4 Resolvase1

### 1.1 Introduction

G-Quadruplexes are highly stable nucleic acid supersecondary structures that form within guanine-rich regions of DNA and RNA. These structures are stabilized by Hoogsteen-bonding interactions, coordinate bonding within the central cavity with monovalent cations, and Van der Waals forces that confer extreme thermal stability to G-quadruplexes (QIN AND HURLEY 2008; STEGLE *et al.* 2009). Recent bioinformatics analysis predicts that there are ~750,000 locations in the human genome likely to form G-quadruplexes (BEDRAT *et al.* 2016). G-Quadruplex forming sequences are not randomly distributed, but evolutionarily conserved and enriched within transcribed regions. Upwards of 40% of human gene promoters containing G-Quadruplex forming motifs (HUPPERT AND BALASUBRAMANIAN 2007). Genes with related functions have been found to have similar degrees of enrichment, with proto-oncogenes and developmental genes having significantly greater G-quadruplex motif enrichment than tumor suppressor genes (EDDY AND MAIZELS 2006; EDDY *et al.* 2011).

With extreme thermal stability, prevalence throughout the human genome, and the ability to affect major regulatory pathways, it is not surprising that there is

an enzymatic mechanism for regulating G-quadruplexes. The enzyme responsible for the majority of G-quadruplex resolving activity in human cells is G4 Resolvase1 (G4R1; alias RHAU and DHX36) (VAUGHN *et al.* 2005). This protein has been found to unwind both tetramolecular and unimolecular G-Quadruplexes in DNA and RNA, while also having the highest reported binding affinity for all know G-Quadruplex binding proteins (CREACY *et al.* 2008; GIRI *et al.* 2011; BOOY *et al.* 2012). The G-Quadruplex resolving activity of G4R1 has been found to be involved in many cellular processes, including telomere biology, transcription and splicing regulation, development, lineage commitment, immune regulation, and the localization of cellular machinery (TRAN *et al.* 2004; KIM *et al.* 2010; LATTMANN *et al.* 2011; SEXTON AND COLLINS 2011; ZHANG *et al.* 2011; BOOY *et al.* 2012; HUANG *et al.* 2012; LAI *et al.* 2012; BICKER *et al.* 2013; BOOY *et al.* 2014; YOO *et al.* 2014; SMALDINO *et al.* 2015; JING *et al.* 2017; NEWMAN *et al.* 2017).

Here, I demonstrate an expression and purification scheme that utilizes the ATP-dependent G-Quadruplex resolving activity of G4R1 to selectively purify catalytically active recombinant G4R1 (rG4R1), combined with a traditional histidine-based purification method (ROUTH *et al.* 2017). This method of purification will allow for advanced biochemical assays of rG4R1, particularly with regards to interactions between rG4R1 and different cellular structures.

## 1.2 Materials and Methods

### 1.2.1 Preparation of G4-DNA Structures for rG4R1 Purification

Formation of Biotinylated G4-DNA G-Quadruplex to be Used for the Purification of rG4R1

To prepare G4-DNA structures for rG4R1 purification, the Z33-Bio oligomer was designed (Table 1). The 3' moiety of biotin is crucial for the purification protocol. Pellet 1 µmole of the oligomer using centrifugation and then resuspend the pellet in 250 µL of water, for a final concentration of 4 mM Z33-Bio. Add 25 µL of G4 buffer (450 mM Tris-HCl, 25 mM EDTA, 2,500 mM NaCl, pH 8) to the solution and mix by pipetting. Then incubate the oligomer solution at 50 °C for 48 hours.

| Oligomer Names | Oligomer Sequences | | |
|---|---|---|---|
| Z33-Bio | 5' | AAA GTG ATG GTG GTG GGG GAA GGA TTC GGA CCT-Biotin | 3' |
| Z33-Alexa$_{680}$ | 5' | Alexa$_{680}$-AAA GTG ATG GTG GTG GGG GAA GGA TTC GGA CCT | 3' |

**Table 1: Oligomer Sequences Used.** These oligomers are based off of Z33, a deoxyoligonucleotide previously characterized to readily form tetramolecular G-quadruplexes (VAUGHN *et al.* 2005).

After the incubation, collect condensation by spinning down the tube (1,000 rcf, 10 seconds). Add 50 µL of a high-mass, hydrophilic polysaccharide (30% Ficoll in $H_2O$) and mix by pipetting.

Pour a 10% acrylamide/1x Tris/Borate/EDTA (TBE) /10% glycerol gel (gel dimensions: 16 cm x 16 cm x 1 mm). After polymerization, wash the wells with 1x TBE and load the processed Z33-Bio oligomer evenly across most of the gel (sample lane loading can be visualized with Schlieren lines). In one lane, add a

small portion of the oligomer that has been heated at 98 °C for 10 minutes to serve as an unstructured control. Add gel loading dye to an empty lane to visually identify migration of sample through the gel. Using 1x TBE, run the gel at 120 V until the loading dye has traversed approximately 1/3 of the gel.

After running the gel, remove one of the glass plates and use ultraviolet (UV) shadowing on the gel (long wavelength: 365 nm) to visualize Z33-Bio bands. Using a clean, DNAse free razor, cut out the bands containing G-Quadruplex-Z33-Bio, which will have run higher up on the gel than the non-quadruplex bands. Place the gel containing the formed G-Quadruplexes in a 50 mL tube and add soaking buffer (1/3 10x TBE, 1/3 saturated sodium acetate pH 8.8, and 1/3 molecular biology grade water) to cover the gel. Incubate the gel at 37 °C overnight.

Transfer the solution, without the gel, to a 15 mL tube and add 1.2x volume isopropanol and 1/10 volume glycogen (glycogen stock at 20 mg/mL in 10 mM Tris-HCl pH 8, 2.5 mM EDTA, 0.05% sodium azide, and pH 8). Mix the solution by pipetting and then incubate the tube at -20 °C for at least 2 hours. Spin the tube at 2,700 rcf in a tabletop centrifuge cooled to 4 °C for 12 min. Decant the supernatant solution and wash the pellet thrice with 50 mM NaCl in 70% EtOH, repelleting the G-quadruplex by centrifugation as described above. Rehydrate the washed pellet by incubating at 4 °C for at least 2 hours. Then, resuspended the pellet in 50 µL TNE buffer (10 mM Tris-HCl, 50 mM NaCl, 0.05 mM EDTA, and pH 8). Quantify the formed G-Quadruplex by adding 5 µL of this solution into 495 µL of water and use a spectrophotometer that allows extinction

coefficient to be determined by entering an oligomer's nucleotide composition (the extinction coefficient of the Z33 DNA oligomer with a base composition of A = 8, C = 3, G = 15, and T = 7 is 341,946 L/mol/cm). Use a dilution factor of 25 (not 100), as the formed G-quadruplex consists of 4 strands of DNA. Aliquot the volume equivalent of 3 ODs ($OD_{260}$ units) per tube and store at -20 °C; for example, if the 5 µL that was added to the 495 µL ofH2O gives an OD260 reading of 3, then prepare 5 µL aliquots.

### 1.2.2 Preparation of G4-DNA Structures for rG4R1 Enzymatic Assays

To prepare G4-DNA structures for rG4R1 enzymatic assay, the Z33-$Alexa_{680}$ oligomer was designed (Table 1). 5' moiety for the $Alexa_{680}$ tag is crucial.

Follow the same procedure as outlined for forming Z33-Bio G-quadruplexes, but do not use UV shadowing, as the Z33-$Alexa_{680}$ bands will be visible to the naked eye. After quantifying the $Alexa_{680}$-labeled DNA G-Quadruplex, dilute the G4-DNA to 0.2 pmol/µL with TNE buffer. Then, add glycerol to a 10% final concentration and store at -20 °C for further use.

### 1.2.3 Production of rG4R1 in a Prokaryotic System

Allow 20 µL of (DE3) PlysS Competent bacteria to thaw on ice while warming 80 µL of SOC medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose) at 37 °C. Once the bacteria have thawed, add 1 µg of pTriEx4-DHX36 plasmid to the bacteria and swirl using a pipette tip to mix (VAUGHN *et al.* 2005). After 5 minutes of incubating the bacteria and plasmid on ice, heat shock the cells at 42 °C for 30

seconds. Place the cells on ice for 2 minutes and then add 80 µL pre-warmed of SOC medium. Shake the cells for 1 hour at 225 rpm and 37 °C to allow for transformed cells to express their antibiotic resistance genes. Spread-plate the transformed bacteria on pre-warmed selection plates containing carbenicillin (CARB) and chloramphenicol (CAM) at 50 µg/mL and 35 µg/mL, respectively and incubate overnight at 37 °C. Inoculate 5 mL of LB media + CARB/CAM in a vented tube with one colony of bacteria from the agar plates (using a p1000 pipette tip), and then incubate the bacteria overnight at 37 °C, shaking at 225 rpm.

Add 1 mL of the liquid culture to 500 mL of Difco™ Terrific Broth (Thermo Fisher, MA, USA) with CARB/CAM in a large baffled flask. After sufficiently swirling the cultures for aeration, incubate the cells at 37 °C without shaking overnight. The next morning, begin shaking the cells at 225 rpm (while still at 37 °C) until the $OD_{600}$ is 0.4 – 0.6. Once the proper bacterial density has been reached, quickly cool the flask to 10 °C on ice. Once the media has reached 10 °C, add IPTG to a 1 mM final concentration (120 mg / 500 mL culture) and incubate the cells at 14 °C while shaking at 80 rpm for 17 – 18 hours. After this, the bacterial concentration will ideally be at 0.8 – 1.2 $OD_{600}$.

Transfer the bacteria to a 500 mL centrifuge tube and centrifuge the bacteria at 3,840 rcf for 20 min at 4 °C. Decant the clarified broth and store the bacterial pellets at -80 °C until ready for protein purification.

### 1.2.4 Purification of Human rG4R1

#### 1.2.4.1 Thaw and Lyse Bacterial Pellets

Thaw previously prepared bacterial pellets in 3 mL of TN buffer (100 mM Tris, 50 mM NaCl, pH 7.5) at room temperature. Once thawed and relatively evenly suspended, bring the volume up to 5 mL with TN buffer. Add lysis buffer (20 mg of lysozyme in 250 µL water) to the resuspended bacteria, and then swirl the bottle in hand for 5 – 10 minutes while the lysozyme digests the bacteria. Add 250 µL of protease inhibitor cocktail (PIC) and 10 µL leupeptin to the bottle, then mix thoroughly. Place the bottles on ice and add 10 mL of cold TN buffer and 22 µL β-mercaptoethanol (BME) before transferring the solution to a 50 mL tube.

In the 50 mL tube, sonicate the bacteria on ice with a digital Sonicator set to 30% amplitude with 2 second alternating on/off cycles for 1 minute (30 seconds total sonication time). Repeat the sonication process for a total of 3 times, with 2 minute breaks between each sonication step.

Add 15 mL of cold 4x saline-sodium citrate (SSC) buffer, 20 µL BME, 50 µL of PIC, and 10 µL of leupeptin to the lysate and mix. After mixing, spin down the lysate in a pre-cooled 4 °C tabletop centrifuge at 23,000 rcf for 20 minutes. Transfer the supernatant liquid to a new 50 mL tube on ice.

#### 1.2.4.2 Prepare G4-Magnetic Beads

Add 1 mL of streptavidin paramagnetic bead (SPB) suspension to a 1.5 mL microcentrifuge tube and pellet the beads with a magnet. After removing the supernatant liquid, wash twice with 2x SSC buffer + 5 mM EDTA pH 8. Resuspend the SPB in another 200 µL of SSC Buffer + EDTA.

Add one 3 OD aliquot of G4-Z33-Bio prepared previously to the SPB suspension and mix by pipetting. Place the microcentrifuge tube on a rotator and rotate at room temperature for 30 – 60 minutes. After rotating, add 1 mL of 0.4% lactalbumin (diluted from 4% lactalbumin (in 2 M glycine pH 7.5 with 0.05% sodium azide and 1x PIC) with 1x Tris-glycine) to block the G4-bound SPB.

**1.2.4.3 Bind Histidine-tagged rG4R1 to Cobalt Beads**

Add 1 mL of cobalt bead (CB) slurry (0.5 mL bead volume) to the 50 mL tube of clarified bacterial lysate, and then incubate for 20 minutes at room temperature on a rotator. After rotating, pellet the CB at 110 rcf for 5 minutes in a pre-cooled 4 °C tabletop centrifuge, and then aspirate the supernatant liquid, leaving ~2 mL of liquid to leave the pellet undisturbed. Wash the pellet with 10 – 15 mL of cold 4x SSC + BME (0.5 µL BME / mL of 4x SSC), but do not pipette the solution to mix. After spinning down the CB pellet, aspirate the liquid and add a second clarified lysate to the pelleted CB, followed by a 20-minute incubation on the rotator at room temperature. Wash the CB twice with 10-15 mL cold 4x SSC + BME, leaving ~2 mL of beads/supernatant liquid in the tube after the second wash. Use a "wide-bore" pipette tip (p1000 pipette with end cut of by a fresh razor blade) to prevent shearing and transfer the CB to a pre-cooled 2 mL tube. Briefly centrifuge the CB at high speed (~18,000 rcf) in a microcentrifuge at 4 °C for 5 - 10 seconds. Gently remove as much supernatant solution as possible without disturbing the protein-bound CB.

#### 1.2.4.4 Elute rG4R1 from Cobalt Beads

Add 0.5 mL histidine elution buffer (HEB) (0.7 M L-histidine, pH 6) to the protein-bound CB without mixing and rotate for 5 minutes at 4 °C. Centrifuge at 18,000 rcf in 4 °C for 1 minute. Transfer the supernatant solution to a 15 mL tube kept on ice, being careful to leave some solution on top of the CB pellet. Repeat these HEB elution steps 2 more times for a total of 3 HEB elutions.

Elute protein using 0.5 mL of 0.2 M EDTA pH 6.0 according to the steps above. Use a gel-loading tip on the end of a 1 mL pipette tip, plunged to the bottom of the CB tube, to obtain as much protein-containing elution buffer as possible.

#### 1.2.4.5 Bind rG4R1 to G4-bound SPB and Elute in an ATP-dependent Manner

Prepare 5x Res buffer (250 mM Tris-Acetate pH 7.8, 250 mM NaCl, 2.5 mM $MgCl_2$, and 50% glycerol) and 3x Res buffer (0.915 mL water, 2 mL 5x Res buffer, 0.33 mL 0.4% lactalbumin, 10 µL BME, 15 µL 1 M $MgCl_2$, 50 µL PIC, and 10 µL leupeptin). Pellet the prepared G4-SPB to the side of the tube with a magnet, discard the supernatant, and add 1 mL of 3x Res buffer, pipetting to mix. Add this G4-SPB solution to the 15 mL tube of HEB/EDTA eluate, then incubate for 15 minutes in a 37 °C water bath, occasionally agitating so the beads do not settle.

On ice, pellet the protein-bound G4-SPB to the side of the tube with a magnet and remove the supernatant liquid. Wash twice with 1 mL of 4x SSC + 0.4% lactalbumin + 0.5 µL/mL BME. During this step, ensure that all of the

magnetic beads have pelleted between washes. Wash the G4-SPB once with 1 mL of 3x Res Buffer.

Pre-warm 1 mL of elution buffer (EB) (0.5 mL 3x Res Buffer, 0.1 mL 0.1 M ATP, and 0.4 mL water) in PCR tubes in a heating block of a thermal cycler set to 37 °C. Pellet G4-SPB with the magnet and suspend the beads in 100 µL EB and immediately transfer the beads to a pre-warmed PCR tube, incubating at 37 °C for 30 seconds. Promptly add 12 µL of 5 M NaCl and pipette vigorously 20-30 times. During this step, minimize the formation of bubbles, as they can denature proteins. Immediately pellet the G4-SPB with a magnet and transfer the protein-containing eluate to a fresh PCR tube on ice, then repeat the 100 µL EB elution process again, producing ~200 µL of EB containing highly pure and catalytically active rG4R1. Set aside two 7 µL aliquots for future enzymatic activity assays, and then store the purified rG4R1 at -80 °C.

### 1.2.5 Quality Control Enzymatic Activity

Prepare 300 µL resolvase assay buffer (RAB) (0.1 mL 3x Res Buffer, 0.01 mL 0.2 pmol/µL G4-Z33-Alexa$_{680}$, 0.03 mL 0.1 M ATP, and 0.16 mL water) and, in a strip of 6 PCR tubes on ice, pipette 40 µL RAB into the first tube and 30 µL RAB into every other tube.

Retrieve a 7 µL aliquot of rG4R1, produced during the purification protocol above, and pipette up and down to mix the protein. Transfer 5 µL rG4R1 to the first tube (already containing 40 µL RAB). Serially transfer 10 µL from the rG4R1-RAB tube to the remaining PCR tubes containing RAB. Incubate this strip of PCR tubes for 30 minutes at 37 °C in a thermocycler, followed by a 4 °C hold. During

the hold, add 2 µL of 0.5 M EDTA pH 8 to each tube (stopping enzymatic activity).

Pour a 10% acrylamide/1x TBE/10% glycerol gel (gel dimensions: 16 cm x 16 cm x 1 mm). After polymerization, wash the wells with 1x TBE. Add 5 µL of 30% Ficoll in $H_2O$ to each tube and load 15 µL from each tube. As a control for G4-DNA mobility, add 4 µL of 30% Ficoll in $H_2O$ to 20 µL RAB and load 15 µL. As a control for unwound monomeric Z33, head 20 µL RAB at 98 °C for 10 minutes, add µL of 30% Ficoll in $H_2O$, and load 15 µL. Using 1x TBE as a running buffer, run the gel at 120 V for ~2 hours. Image the gel, looking for $Alexa_{680}$ bands. Highly active preparations of rG4R1 will yield complete resolution of $Alexa_{680}$-labeled Z33 in the first lanes that were loaded on the gel, as shown in Figure 1.


**Figure 1: Quality-control rG4R1 Enzymatic Activity Assay.** Lanes 1 – 6: A constant concentration of TAMRA-labeled tetramolecular G4-DNA was incubated at 37 °C for 30 min in the presence of 4x serial dilutions of purified rG4R1 representing 3.9 µL, 0.83 µL, 0.2 µL, 0.05 µL, 0.013 µL, and 0.003 µL, respectively. Lane 7: Tetramolecular G4-DNA in the absence of rG4R1. Lane 8: Tetramolecular G4-DNA boiled to reduce the G4 structure into monomers in the absence of rG4R1.

### 1.2.6 Pooling of Highly Active rG4R1 Preparations, Aliquoting, and Storage

Calculate the total volume of rG4R1 that will be aliquoted (7 µL is common for downstream assays) and fill the block of a thermal cycler set to 4 °C with strips of PCR tubes. Retrieve PCR tubes containing highly active preparations of rG4R1, thaw in hand, and place on ice when there is a small amount of ice remining in the tube. Combine all of these preparations into a pre-chilled 15 mL tube, mixing well while minimizing bubble formation.

Use an automatic repeating pipette to dispense the desired aliquot volume into the pre-chilled PCR tubes in the thermal cycler. As soon as 1 – 2 strips are completed, a second person should close the lids and transfer the strips to 96-well wafers floating in liquid nitrogen, in order to preserve enzymatic activity. After all of the strips have been flash-frozen, transfer the tubes from liquid nitrogen to dry ice, and then into a -80 °C freezer for long-term storage.

### 1.2.7 Quantification of Purified rG4R1

Pour a 5% stacking/12% resolving acrylamide/SDS gel for protein separation. Thaw ~50 µL of aliquoted rG4R1, combine into one tube, and mix well. Confirm the exact volume with a pipette and add an equal amount of 2x Laemmli sample buffer (65.8 mM Tris-HCl, 26.3% (w/v) glycerol, 2.1% SDS, 0.02% bromophenol blue, pH 6.8) + BME (50 µL/950 µL of 2x Laemmli buffer). Denature the protein by incubating at 98 °C for 10 minutes.

Prepare dilutions of broad range molecular weight markers in 1x Laemmli buffer + BME such that the following protein amounts can be loaded onto the gel

as standards: 31.3, 62.5, 125, 250, and 500 ng. (Each protein standard contained in the broad-range MW markers is present at 0.1 µg/µL, except for the 50 kDa protein, which is present at 0.3 µg/µL.) Wash the wells with 1x SDS/glycine running buffer and add 50 µL and 25 µL denatured rG4R1 solution (25 and 12.5 µL protein). Load the protein standards prepared, being careful to be as precise as possible to ensure accurate quantification. Run the gel at 120 V until the dye has traversed ~2/3 of the gel.

Remove the portion of the gel not containing any protein and place the protein-containing portion in a glass dish. Stain the gel with Coomassie dye (50 mg of Coomassie R-250 per 100 mL of 50:10:40 v/v methanol:acetic acid:$H_2O$ that has been filtered through a filter-paper funnel) overnight at room temperature on an orbital shaker to ensure adequate agitation of the gel.

De-stain the gel with 30:10:60 v/v methanol:acetic acid:$H_2O$, using balled up KimWipes™ (Kimberly-Clark, TX, USA) to expedite the de-staining process. During this process, the de-staining solution may need to be replaced. Continue de-staining until the background is low enough to visualize rG4R1 and the protein standards. A gel at this stage is shown in Figure 2.

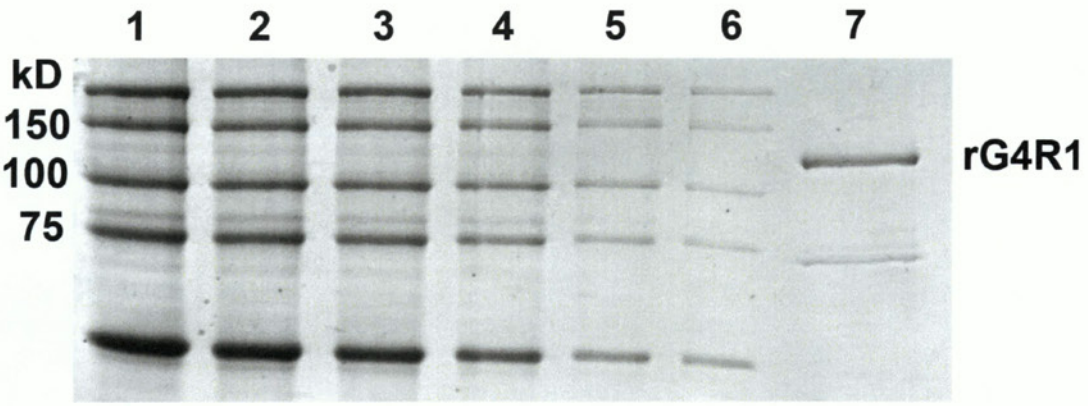


**Figure 2: Quantification of Purified rG4R1.** Lanes 1 – 6: Broad-range molecular weight protein markers were loaded in these quantities: 500, 250, 125, 62.5, 31.3, and 15.7 ng, respectively, and were used to generate a standard curve of protein concentrations. Lane 7: 35 µL of rG4R1. This particular gel was quantified, as part of a triplicate set of gels, resulting in an average protein concentration of 62 ± 22 nM standard deviation (n = 9).

Place the de-stained gel in a sheet protector and scan the gel at ≥ 300 dpi resolution. Within an image analysis program, prepare standard curves from the protein standards that correspond to 75, 100, and 150 kDa (not including the background). From these standard curves, the concentration of rG4R1 can be interpolated. From the calculated concentration and known volume added, determine the mass of rG4R1 added. Then calculate the molar amount using rG4R1's molecular weight (120,000 g/mol). This will produce 3 calculated molar values for rG4R1. Repeating this procedure 2 more times will produce 9 estimates, which are then to be averaged into a single batch-specific rG4R1 concentration, typically in the range of 20 – 100 nM.

## 1.3 Discussion

This protocol demonstrates a highly efficient expression, purification, quantification, and enzymatic assay scheme for the isolation of G4R1, the protein product of *DHX36*. This protocol implements two purification steps: traditional His-tag affinity purification with cobalt beads and enzymatic purification with G4-DNA-conjugated paramagnetic beads. The second purification process is key to the great potential of this purification. Exposing an already semi-pure eluate to G4-DNA-conjugated SPB will allow for the recovery of almost all rG4R1 proteins, due to their extremely high binding affinity for G-quadruplexes; almost all the impurities in the eluate without G-quadruplex affinities will be removed at this step, with non-specific G-Quadruplex binding proteins being washed off of the SPB. Then, the catalytic activity-based purification, adding in the necessary molecules (ATP and $MgCl_2$) for rG4R1 to resolve G-Quadruplexes, allows for the elimination of any malfunctioning rG4R1 proteins that bind G-quadruplexes without possessing G-quadruplex resolving capabilities. The production of highly pure and exclusively catalytically active rG4R1 is a great tool for biochemical assays. This selectivity for catalytic activity separates this protocol from other protein purification methods, which are unable to separate enzymatically active proteins from enzymatically dead proteins (Booy *et al.* 2015).

# CHAPTER TWO

# G-Quadruplexes in C9orf72 ALS

## 2.1 Introduction

A recent breakthrough in the field of research for amyotrophic lateral sclerosis (ALS) was made when it was discovered that many patients suffering from ALS had a similar genetic mutation. This mutation is a hexanucleotide repeat expansion of the sequence GGGGCC within the gene *C9orf72* (DeJesus-Hernandez *et al.* 2011; Renton *et al.* 2011). Within healthy individuals, this repeat expansion is repeated less than 30 times. People suffering from this variant of ALS typically have more than 30 copies of this guanine-rich sequence, with higher repeat numbers correlating with earlier onset and increased severity of the disease. Due to the high local concentration of guanine, this sequence has been found to form G-quadruplexes (Donnelly *et al.* 2013; Haeusler *et al.* 2014).

G-quadruplex RNA sequences, forming due to the transcription of the repeat expansions, are commonly identified features of ALS and are considered to generate cellular toxicity by sequestering crucial RNA binding proteins (Mizielinska *et al.* 2013; Reddy *et al.* 2013; Zu *et al.* 2013; Haeusler *et al.* 2014; Conlon *et al.* 2016). Because the repeat expansion is 6 nucleotides long, the

expansion, when translated, creates a glycine/arginine dipeptide repeat that sequesters cellular proteasomes, interfering with protein degradation pathways (GUPTA *et al.* 2017; GUO *et al.* 2018).

It has previously been shown that G4R1 is responsible for the majority of all G-quadruplex resolving activity in human cells (VAUGHN *et al.* 2005; CREACY *et al.* 2008). It is likely that G4R1 is functionally involved in ALS pathology because of G4R1 having the highest G-quadruplex binding affinity of any know G-quadruplex binding protein, G4R1 having specificity for both DNA and RNA, allowing it to interact with the guanine-rich repeat expansion at multiple levels, and G4R1 has been found to associated with another ALS-linked protein, TDP-43 (CREACY *et al.* 2008; FREIBAUM *et al.* 2010; LATTMANN *et al.* 2010; GIRI *et al.* 2011; SMALDINO *et al.* 2015).

Furthermore, FUS, a protein known to cause ALS when mutated, possesses a G-quadruplex binding domain (TAKAHAMA *et al.* 2015; COPPEDE *et al.* 2018). TDP-43, a protein known to associate with G4R1 and contribute to ALS when dysfunctional, utilizes RNA G-quadruplex affinity to transport mRNAs into neurites (FREIBAUM *et al.* 2010; ISHIGURO *et al.* 2016). These intersections of ALS pathology and G-quadruplex biology outside of the *C9orf72* repeat expansion might be due to a deeper entanglement of ALS with G-Quadruplexes. If so, G4R1, the foremost human G-quadruplex helicase, will likely be involved because of the reasons mentioned above.

## 2.2 C9 ALS Working Model

Due to G4R1's affinity for both DNA and RNA G-quadruplexes, there is great potential for G4R1 to be involved in ALS pathology at multiple levels. Our current working model for how G4R1 interacts with ALS pathology is that, initially, the guanine-rich expansion repeat will generate a high number of localized G-quadruplexes. All of these G-quadruplexes would cause G4R1 proteins, with their particularly high quadruplex affinities, to bind the expansion repeat region. Normally, G-quadruplexes would prevent transcription by impeding the transcriptional machinery (Figure 3-A). However, the localization of G4R1 would allow for the unwinding of *C9orf72*, allowing for transcription to occur (Figure 3-B). These transcripts, containing the guanine-rich expansion repeat, will have the same potential to form G-quadruplexes. However, there is the potential for there to be many more transcripts, possibly producing many G-quadruplexes in the cell that have the potential to sequester RNA binding proteins that are crucial to normal cellular function (Figure 3-C). However, the rapid abundance of G-quadruplexes may trigger the cell to increase the expression of G4R1, which could cause problems in multiple ways. The G4R1 may localize to the original repeat expansion, allowing for even more transcription to occur, creating a positive feedback loop (Figure 3-D). Additionally, the expressed G4R1 may resolve the G-quadruplex RNA, allowing it to be translated (Figure 3-E). As mentioned above, this would produce dysfunctional proteins that would serve to hamper the cell's protein degradation system by inhibiting proteasome function (Gupta *et al.* 2017; Guo *et al.* 2018).

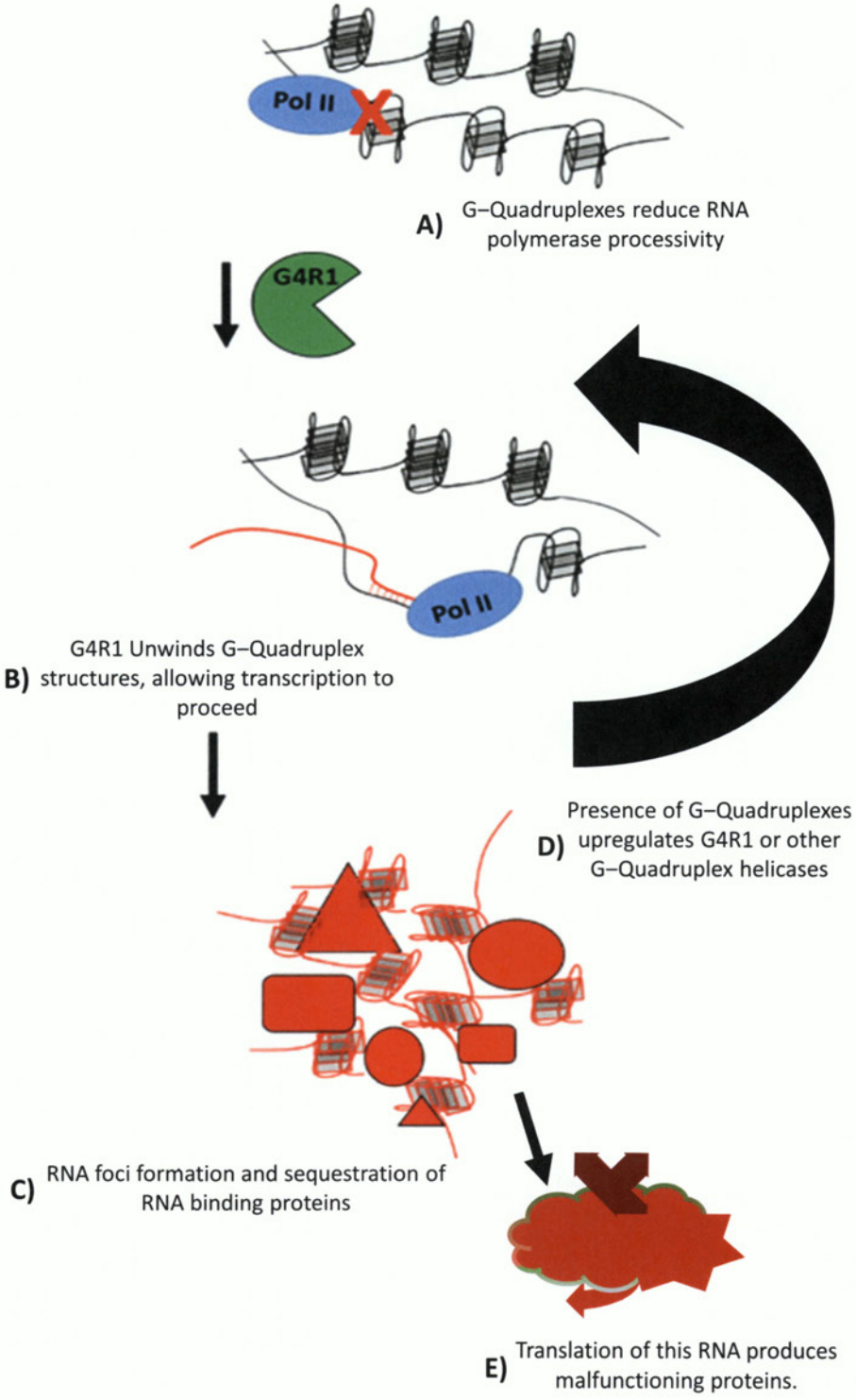


**Figure 3: Working Model of G4R1's Involvement in ALS Pathology**

## 2.3 Future Research

Due to technical issues, we were unable to transfect HeLa cells with plasmids containing the *C9orf72* expansion repeat. Once a protocol for transfecting HeLa cells with these guanine rich plasmids is optimized, our lab will begin several experiments regarding G4R1. Chromatin immunoprecipitation (ChIP) will be used to verify an initial ChIP assay suggesting that G4R1 localizes to the expansion repeat localization. Due to G-quadruplexes inhibiting transcription and G4R1 knockdowns having been previously shown to decrease transcription of a guanine-rich gene, we hypothesize that knockdown of G4R1 with siRNA will decrease transcription of the *C9orf72* expansion repeat, thereby reducing *C9* mRNA and *C9* RNA foci levels, which will be quantified by qPCR and RNA FISH, respectively. Utilizing plasmids with different repeat numbers in HeLa cells, we will determine if greater expansion repeat size correlates with increased G4R1 levels with quantitative western blots.

Due to TDP-43's known association with RNA G-quadruplexes, I propose that TDP-43's bound RNA sequences be identified by RNA immunoprecipitation (RIP). The usage of RIP allows for the quantification off TDP-43 bound to *C9orf72* transcripts, relative to the number of repeat expansions. The benefit of this technique is that it can readily be transitioned to quantify the interactions between TDP-43 and G4R1 through co-immunoprecipitation (co-IP).

The usage of co-IP would allow for the identification of possible interactions between FUS and G4R1. If these interactions are identified, it is possible that these will be due to the association of FUS with G-quadruplexes in

both DNA and RNA, so the usage of RIP and chromatin immunoprecipitation would identify if FUS was bound to *C9orf72* or *C9orf72* transcripts.

## LITERATURE CITED


Bedrat, A., L. Lacroix and J.-L. Mergny, 2016 Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Research 44: 1746-1759.

Bicker, S., S. Khudayberdiev, K. Weiß, K. Zocher, S. Baumeister *et al.*, 2013 The DEAH-box helicase DHX36 mediates dendritic localization of the neuronal precursor-microRNA-134. Genes & Development 27: 991-996.

Booy, E. P., R. Howard, O. Marushchak, E. O. Ariyo, M. Meier *et al.*, 2014 The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. Nucleic Acids Res 42: 3346-3361.

Booy, E. P., E. K. McRae and S. A. McKenna, 2015 Biochemical characterization of G4 quadruplex telomerase RNA unwinding by the RNA helicase RHAU. Methods Mol Biol 1259: 125-135.

Booy, E. P., M. Meier, N. Okun, S. K. Novakowski, S. Xiong *et al.*, 2012 The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. Nucleic Acids Research 40: 4110-4124.

Conlon, E. G., L. Lu, A. Sharma, T. Yamazaki, T. Tang *et al.*, 2016 The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. Elife 5.

Coppede, F., A. Stoccoro, L. Mosca, R. Gallo, C. Tarlarini *et al.*, 2018 Increase in DNA methylation in patients with amyotrophic lateral sclerosis carriers of not fully penetrant SOD1 mutations. Amyotroph Lateral Scler Frontotemporal Degener 19: 93-101.

Creacy, S. D., E. D. Routh, F. Iwamoto, Y. Nagamine, S. A. Akman *et al.*, 2008 G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates. The Journal of Biological Chemistry 283: 34626-34634.

DeJesus-Hernandez, M., Ian R. Mackenzie, Bradley F. Boeve, Adam L. Boxer, M. Baker *et al.*, 2011 Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron 72: 245-256.

Donnelly, C. J., P. W. Zhang, J. T. Pham, A. R. Haeusler, N. A. Mistry *et al.*, 2013 RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron 80: 415-428.

Eddy, J., and N. Maizels, 2006 Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Research 34: 3887-3896.

Eddy, J., A. C. Vallur, S. Varma, H. Liu, W. C. Reinhold *et al.*, 2011 G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. Nucleic Acids Research 39: 4975-4983.

Freibaum, B. D., R. Chitta, A. A. High and J. P. Taylor, 2010 Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. Journal of proteome research 9: 1104-1120.

Giri, B., P. J. Smaldino, R. G. Thys, S. D. Creacy, E. D. Routh *et al.*, 2011 G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA. Nucleic Acids Res 39: 7161-7178.

Guo, Q., C. Lehmer, A. Martinez-Sanchez, T. Rudack, F. Beck *et al.*, 2018 In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment. Cell 172: 696-705 e612.

Gupta, R., M. Lan, J. Mojsilovic-Petrovic, W. H. Choi, N. Safren *et al.*, 2017 The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome. eNeuro 4: ENEURO.0249-0216.2017.

Haeusler, A. R., C. J. Donnelly, G. Periz, E. A. J. Simko, P. G. Shaw *et al.*, 2014 C9orf72 Nucleotide Repeat Structures Initiate Molecular Cascades of Disease. Nature 507: 195-200.

Huang, W., P. J. Smaldino, Q. Zhang, L. D. Miller, P. Cao *et al.*, 2012 Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1. Nucleic Acids Research 40: 1033-1049.

Huppert, J. L., and S. Balasubramanian, 2007 G-quadruplexes in promoters throughout the human genome. Nucleic Acids Research 35: 406-413.

Ishiguro, A., N. Kimura, Y. Watanabe, S. Watanabe and A. Ishihama, 2016 TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. Genes Cells 21: 466-481.

Jing, H., Y. Zhou, L. Fang, Z. Ding, D. Wang *et al.*, 2017 DExD/H-Box Helicase 36 Signaling via Myeloid Differentiation Primary Response Gene 88 Contributes to NF-κB Activation to Type 2 Porcine Reproductive and Respiratory Syndrome Virus Infection. Frontiers in Immunology 8: 1365.

Kim, T., S. Pazhoor, M. Bao, Z. Zhang, S. Hanabuchi *et al.*, 2010 Aspartate-glutamate-alanine-histidine box motif (DEAH)/RNA helicase A helicases sense microbial DNA in human plasmacytoid dendritic cells. Proceedings of the National Academy of Sciences of the United States of America 107: 15181-15186.

Lai, J. C., S. Ponti, D. Pan, H. Kohler, R. C. Skoda *et al.*, 2012 The DEAH-box helicase RHAU is an essential gene and critical for mouse hematopoiesis. Blood 119: 4291-4300.

Lattmann, S., B. Giri, J. P. Vaughn, S. A. Akman and Y. Nagamine, 2010 Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. Nucleic Acids Research 38: 6219-6233.

Lattmann, S., M. B. Stadler, J. P. Vaughn, S. A. Akman and Y. Nagamine, 2011 The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. Nucleic Acids Research 39: 9390-9404.

Mizielinska, S., T. Lashley, F. E. Norona, E. L. Clayton, C. E. Ridler *et al.*, 2013 C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathologica 126: 845-857.

Newman, M., R. Sfaxi, A. Saha, D. Monchaud, M.-P. Teulade-Fichou *et al.*, 2017 The G-Quadruplex-Specific RNA Helicase DHX36 Regulates p53 Pre-mRNA 3'-End Processing Following UV-Induced DNA Damage. Journal of Molecular Biology 429: 3121-3131.

Qin, Y., and L. H. Hurley, 2008 Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie 90: 1149-1171.

Reddy, K., B. Zamiri, S. Y. R. Stanley, R. B. Macgregor and C. E. Pearson, 2013 The Disease-associated r(GGGGCC)(n) Repeat from the C9orf72 Gene Forms Tract Length-dependent Uni- and Multimolecular RNA G-quadruplex Structures. The Journal of Biological Chemistry 288: 9860-9866.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson *et al.*, 2011 A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72: 257-268.

Routh, E. D., S. D. Creacy, P. E. Beerbower, S. A. Akman, J. P. Vaughn *et al.*, 2017 A G-quadruplex DNA-affinity Approach for Purification of Enzymatically Active G4 Resolvase1. J Vis Exp: e55496.

Sexton, A. N., and K. Collins, 2011 The 5' Guanosine Tracts of Human Telomerase RNA Are Recognized by the G-Quadruplex Binding Domain of the RNA Helicase DHX36 and Function To Increase RNA Accumulation. Molecular and Cellular Biology 31: 736-743.

Smaldino, P. J., E. D. Routh, J. H. Kim, B. Giri, S. D. Creacy *et al.*, 2015 Mutational Dissection of Telomeric DNA Binding Requirements of G4 Resolvase 1 Shows that G4-Structure and Certain 3'-Tail Sequences Are Sufficient for Tight and Complete Binding. PLoS One 10: e0132668.

Stegle, O., L. Payet, J.-L. Mergny, D. J. C. MacKay and J. L. Huppert, 2009 Predicting and understanding the stability of G-quadruplexes. Bioinformatics 25: i374-i1382.

Takahama, K., A. Miyawaki, T. Shitara, K. Mitsuya, M. Morikawa *et al.*, 2015 G-Quadruplex DNA- and RNA-Specific-Binding Proteins Engineered from the RGG Domain of TLS/FUS. ACS Chem Biol 10: 2564-2569.

Tran, H., M. Schilling, C. Wirbelauer, D. Hess and Y. Nagamine, 2004 Facilitation of mRNA Deadenylation and Decay by the Exosome-Bound, DExH Protein RHAU. Molecular Cell 13: 101-111.

Vaughn, J. P., S. D. Creacy, E. D. Routh, C. Joyner-Butt, G. S. Jenkins *et al.*, 2005 The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. J Biol Chem 280: 38117-38120.

Yoo, J.-S., K. Takahasi, C. S. Ng, R. Ouda, K. Onomoto *et al.*, 2014 DHX36 Enhances RIG-I Signaling by Facilitating PKR-Mediated Antiviral Stress Granule Formation. PLoS Pathogens 10: e1004012.

Zhang, Z., T. Kim, M. Bao, V. Facchinetti, Sung Y. Jung *et al.*, 2011 DDX1, DDX21, and DHX36 Helicases Form a Complex with the Adaptor Molecule TRIF to Sense dsRNA in Dendritic Cells. Immunity 34: 866-878.

Zu, T., Y. Liu, M. Banez-Coronel, T. Reid, O. Pletnikova *et al.*, 2013 RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. Proc Natl Acad Sci U S A 110: E4968-4977.